IME 775 — Lecture 4

Linear Systems, Eigenanalysis, and Dimensionality Reduction

---

1. Linear Systems: The Core Problem

Problem: Given $\mathbf{A}$ and $\mathbf{b}$, find $\mathbf{x}$ such that:

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

ML Context:

- $\mathbf{A}$: design matrix (features)
- $\mathbf{x}$: weights to learn
- $\mathbf{b}$: target outputs

Example:

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

$$2x_1 + x_2 = 5$$

$$x_1 + 3x_2 = 11$$

Solution:

$$x_1 = 2$$

$$x_2 = 3$$

---

2. Matrix Inverse

Definition: For square $\mathbf{A}$, the inverse $\mathbf{A}^{-1}$ satisfies:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

Solution to linear system:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

2×2 Inverse Formula:

$$\mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad \text{for } \mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Workout: Find the inverse of $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$:

3. Determinant

Definition: For 2×2 matrix:

$$\det(\mathbf{A}) = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

Key Facts:

- $\det(\mathbf{A}) \neq 0 \Leftrightarrow \mathbf{A}$ is invertible
- $|\det(\mathbf{A})|$ = area scaling factor
- $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$

Workout: Compute $\det \begin{bmatrix} 3 & 2 \\ 6 & 4 \end{bmatrix}$. Is this matrix invertible?

4. Singular Matrices

Definition: A matrix is **singular** if $\det(\mathbf{A}) = 0$.

Equivalent conditions:

- $\mathbf{A}^{-1}$ does not exist
- Rows/columns are linearly dependent
- $\mathbf{A}\mathbf{x} = \mathbf{b}$ has no unique solution
- $\mathbf{A}$ collapses space (loses dimension)

ML Implication: Singular design matrix → model is ill-conditioned, need regularization.

---

5. Over/Under-Determined Systems

Overdetermined ($m > n$): More equations than unknowns.

- Typically no exact solution (noisy data)
- Find **least-squares** solution: minimize $||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2$

Underdetermined ($m < n$): Fewer equations than unknowns.

- Infinitely many solutions
- Find **minimum-norm** solution: smallest $||\mathbf{x}||_2$

Workout: A system has 100 data points and 5 features. Is it over or underdetermined?

---

6. Moore-Penrose Pseudo-Inverse

Definition: $\mathbf{A}^+$ exists for ANY matrix (even non-square, singular).

For overdetermined (full column rank):

$$\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$$

For underdetermined (full row rank):

$$\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$$

Least-squares solution:

$$\mathbf{x}^* = \mathbf{A}^+\mathbf{b}$$

---

7. Normal Equations

For least-squares, solve:

$$(\mathbf{A}^T\mathbf{A})\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

This is the foundation of **linear regression**!

Workout: For data points $(0, 1), (1, 3), (2, 5)$, set up the normal equations for $y = mx + c$:

Solution:

Given data points:

$$(0,1),\ (1,3),\ (2,5)$$

Each row of A is

$$[x_i \ \ 1]$$

, and b contains the

$$y_i\text{'}s$$

:

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} m \\ c \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$ So the overdetermined system is:

$$A\mathbf{x} \approx \mathbf{b}$$

$$A^T = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 13 \\ 9 \end{bmatrix}$$

$$\boxed{\begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 13 \\ 9 \end{bmatrix}}$$

$$\boxed{\begin{aligned} 5m + 3c = 13 \\ 3m + 3c = 9 \end{aligned}}$$

---

8. Eigenvalues and Eigenvectors

Definition: For square $\mathbf{A}$, if:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

then λ is an **eigenvalue** and $\mathbf{v}$ is the corresponding **eigenvector**.

Geometric Meaning: Eigenvectors are directions that only get **scaled** (not rotated) by $\mathbf{A}$.

Workout: Verify that $\mathbf{v} = [1, 1]^T$ is an eigenvector of $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. What is λ?

---

9. Finding Eigenvalues

Characteristic Equation:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

For 2×2: this gives a quadratic in λ.

Workout: Find eigenvalues of $\mathbf{A} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$:

Step 1: Write $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$

Step 2: Expand and solve:

---

10. Finding Eigenvectors

For each eigenvalue λ_i, solve:

$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{v} = \mathbf{0}$$

Workout: For $\mathbf{A} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$ with $\lambda = 5$, find the eigenvector:

---

11. Properties of Eigenvectors

Theorem: Eigenvectors corresponding to **distinct** eigenvalues are linearly independent.

Proof sketch:

Suppose $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 = \mathbf{0}$

Apply $\mathbf{A}$: $\alpha_1\lambda_1\mathbf{v}_1 + \alpha_2\lambda_2\mathbf{v}_2 = \mathbf{0}$

Subtract: $\alpha_2(\lambda_2 - \lambda_1)\mathbf{v}_2 = \mathbf{0}$

Since $\lambda_1 \neq \lambda_2$: $\alpha_2 = 0$. Similarly, $\alpha_1 = 0$. $\square$

---

12. Symmetric Matrices: Special Properties

For **symmetric** $\mathbf{A} = \mathbf{A}^T$:

1. All eigenvalues are **real**
2. Eigenvectors are **orthogonal**
3. $\mathbf{A}$ is always diagonalizable

Why it matters: Covariance matrices are always symmetric!

Workout: Verify $\begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$ is symmetric. Find its eigenvalues:

---

13. Spectral Theorem

Theorem: For symmetric $\mathbf{A}$:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$$

where:

- $\mathbf{Q}$: orthogonal matrix (columns = eigenvectors)
- $\mathbf{\Lambda}$: diagonal matrix (eigenvalues on diagonal)

Also written as:

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{q}_i \mathbf{q}_i^T$$

---

14. Matrix Diagonalization

General form: If $\mathbf{A}$ has n linearly independent eigenvectors:

$$\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}$$

where columns of $\mathbf{S}$ are eigenvectors.

Power application:

$$\mathbf{A}^k = \mathbf{S}\mathbf{\Lambda}^k\mathbf{S}^{-1} = \mathbf{S} \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix} \mathbf{S}^{-1}$$

Workout: If $\lambda_1 = 0.9$ and $\lambda_2 = 0.5$, what happens to $\mathbf{\Lambda}^{100}$?

---

15. Spectral Radius

Definition:

$$\rho(\mathbf{A}) = \max_i |\lambda_i|$$

Significance for ML:

- $\rho(\mathbf{A}) < 1$: powers decay → **stable**
- $\rho(\mathbf{A}) > 1$: powers grow → **unstable**
- $\rho(\mathbf{A}) = 1$: borderline

RNN gradient flow: If weight matrix has $\rho > 1$, gradients explode!

---

16. Orthogonal Matrices

Definition: $\mathbf{Q}$ is orthogonal if:

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$$

Properties:

- $\mathbf{Q}^{-1} = \mathbf{Q}^T$ (inverse is just transpose!)

- $\|\mathbf{Qx}\|_2 = \|\mathbf{x}\|_2$ (preserves length)
- $\det(\mathbf{Q}) = \pm 1$

Examples: Rotation matrices, reflection matrices.

Workout: Verify $\mathbf{Q} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ satisfies $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$:

---

17. Rotation Matrix Eigenvalues

For rotation by angle θ:

$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Eigenvalues: $\lambda = e^{\pm i\theta} = \cos\theta \pm i\sin\theta$

Insight: Complex eigenvalues encode rotation angle!

- Eigenvalue of 1 → axis of rotation (in 3D)

Summary: Key Formulas

Concept	Formula
Matrix Inverse (2×2)	$\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
Characteristic Eqn	$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
Eigenvector Eqn	$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$
Spectral Theorem	$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$
Matrix Power	$\mathbf{A}^k = \mathbf{S}\mathbf{\Lambda}^k\mathbf{S}^{-1}$
Pseudo-inverse	$\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$

References

Math and Architectures of Deep Learning by K. Chaudhury